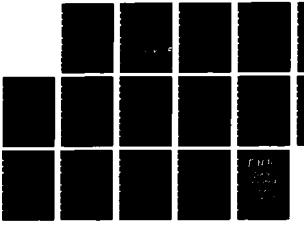
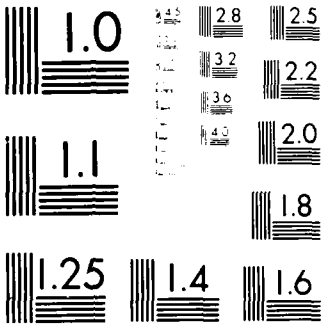
AD-A186 436 SUBLINEAR UPPER BOUNDS FOR STOCHASTIC PROGRAMS WITH RECOURSE REVISION(U) MICHIGAN UNIV ANN ARBOR J R BIRGE ET AL JUN 87 TR-86-26 N00014-86-K-0628 1/1

UNCLASSIFIED F/G 12/4 NL





Sublinear Upper Bounds for Stochastic Programs with Recourse

John R. Birge
Industrial and Operations Engineering
University of Michigan
Ann Arbor, Michigan, USA 48109

Roger J-B Wets
Mathematics
University of California
Davis, California, USA 95616

Technical Report 86-26

Department of Industrial and Operations Engineering

The University of Michigan
College of Engineering
Ann Arbor, Michigan 48109-2117

(12)

Sublinear Upper Bounds for Stochastic Programs with Recourse

John R. Birge
Industrial and Operations Engineering
University of Michigan
Ann Arbor, Michigan, USA 48109

Roger J-B Wets
Mathematics
University of California
Davis, California, USA 95616

Technical Report 86-26

Revised June 1987

Sublinear Upper Bounds for Stochastic Programs with Recourse

John R. Birge*
Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI, USA 48109

Roger J-B. Wets*
Mathematics
University of California
Davis, CA, USA 95616

Abstract: Separable sublinear functions are used to provide upper bounds on the recourse function of a stochastic program. The resulting problem's objective involves the inf-convolution of convex functions. A dual of this problem is formulated to obtain an implementable procedure to calculate the bound. Function evaluations for the resulting convex program only require a small number of single integrations in contrast with previous upper bounds that require a number of function evaluations that grows exponentially in the number of random variables. The sublinear bound can often be used when other suggested upper bounds are intractable. Computational results indicate that the sublinear approximation provides good, efficient bounds on the stochastic program objective value.

Keywords: stochastic programming, sublinear function, simple recourse problem, recourse model, duality, approximation.

Abbreviated title: Sublinear upper bounds.

* This research has been partially supported by the National Science Foundation. The first author's work was also supported in part by Office of Naval Research Grant N00014-86-K-0628 and by the National Research Council under a Research Associateship at the Naval Postgraduate School, Monterey, California.

1. Introduction

We consider the following *stochastic program with recourse*:

$$\text{find } x \in \Re^{n_1}, \chi \in \Re^{m_2} \text{ such that } Ax = b, Tx - \chi = 0, x \geq 0, \tag{1.1}$$

$$\text{and } z = cx + \Psi(\chi) \text{ is minimized,}$$

where $\Psi(\chi) = \mathrm{E}\{\psi(\chi, \xi)\}$ and $\psi(\chi, \xi)$ is the *recourse function* defined by

$$\psi(\chi, \xi) = \inf_{y \in \Re^{n_2}_+} \{qy \mid Wy = \xi - \chi\}. \tag{1.2}$$

The random m_2-vector ξ is defined on a probability space $(\Xi, \mathcal{B}, P)$. The vectors $b \in \Re^{m_1}$, $c \in \Re^{n_1}$, $q \in \Re^{n_2}$, and matrices, $A \in \Re^{m_1 \times n_1}$, $T \in \Re^{m_2 \times n_1}$, $W \in \Re^{m_2 \times n_2}$, are deterministic. In more general models, q and T may be stochastic, but we confine our attention to the case of random right-hand sides in (1.2).

Difficulties in performing the multiple integration to evaluate $\Psi(\chi)$ make the solution of (1.1) especially complex. Solution procedures that do not assume special structure, therefore, involve some approximation to $\Psi(\chi)$ and its derivatives. One type of procedure is to sample from ξ randomly and to use sample information to guide an optimization algorithm. These *stochastic quasi-gradient methods* (see Ermoliev (1983) and Ermoliev and Gaivoronski (1987)) have asympotic convergence properties, but they are limited by a lack of computable bounds.

Other approximation procedures for (1.1) generally rely on discretizations of Ξ (Huang, Ziemba, and Ben-Tal (1977), Kall and Stoyan (1982), Birge and Wets (1986a)). The lower bounds are based on Jensen's inequality and, in general, require a small number of function evaluations. The upper bounds, however, require evaluating $\psi(\chi, \cdot)$ at all extreme points of Ξ (or some region containing Ξ), i.e., at least 2^{m_2} solutions of (1.2) with varying ξ. These approaches are then limited to problems with small m_2. The sublinear upper bound we propose requires only solving $O(m_2)$ linear programs (1.2), so that a much broader class of problems can be considered.

Various methods have been suggested for solving the large–scale linear program associated with a discretization of Ξ (Wets (1987)). These methods include basis factorization

(Strazicky (1980), Kall (1979)), inner linearization (Dantzig and Madansky (1961)), and outer linearization (Van Slyke and Wets (1969)). Implementations are described in Kall and Keller (1983) and Birge (1985, 1987). In general, the number of realizations in the discretization is limited to numbers in the hundreds (i.e., $m_2 < 10$). This discretization level is not always enough to provide close bounds on the objective value. Solution times can also become extremely long. The situation is better, however, if (1.1) has special structure.

The *stochastic program with simple recourse* is a special instance of (1.1) in which $W = (I, -I)$, where I is an $m_2 \times m_2$ identity matrix. This allows for the use of general nonlinear optimization procedures (Nazareth and Wets (1986)) and combinations of linear and nonlinear procedures (Qi (1986)). Separability reduces the m_2–dimensional integration to m_2 independent one–dimensional integrals; it renders possible the direct computation of the values and the derivatives of Ψ for use in optimization.

We obtain efficient upper bounds by replacing the recourse function with functions similar to simple recourse functions. We generalize and extend the *ray function* approximation in Birge and Wets (1986a) and the efficient network implementation in Wallace (1987). We show that linear transformations of the random vectors can be used to obtain a *variety* of separable, sublinear bounding functions. We also give a dual–based solution procedure for combining these bounding functions. The separable sublinear functions allow for monotropic optimization (Rockafellar (1984)). The convex hull of several of these functions is used in the calculation of the upper bound. In our experience, the solution of this upper bounding problem provides quick solutions that are close in value to the optimum. It may be especially valuable in providing initial solution values (as in Birge and Wets (1984)) for further optimization procedures (Nazareth and Wets (1986)).

The simple recourse problem and its properties are described in Section 2. Section 3 describes the sublinear approximation procedure and its relation to the simple recourse problem. Section 4 explains how the sublinear approximation is solved using its dual, and Section 5 discusses our implementation. Section 6 presents computational results. Section 7 discusses possible extensions.

2. The Simple Recourse Problem

The simple recourse problem has a separable recourse function that is written as $\Psi(\chi) = \sum \Psi_i(\chi_i)$ where $\Psi_i(\chi_i) = \int \psi_i(\chi_i, \xi_i) dP_i(\xi_i)$, P_i is the marginal distribution function of ξ_i, which we assume has bounded first and second moments, and

$$\begin{aligned} \psi_i(\chi_i, \xi_i) &= \inf\{q_i^+ y_i^+ + q_i^- y_i^- \mid y_i^+ - y_i^- = \xi_i - \chi_i, y_i^+ \geq 0, y_i^- \geq 0\} \\ &= \begin{cases} q_i^+(\xi_i - \chi_i) & \text{if } \chi_i \leq \xi_i, \\ q_i^-(\chi_i - \xi_i) & \text{if } \chi_i > \xi_i. \end{cases} \end{aligned} \tag{2.1}$$

Detailed properties of the simple recourse problem are given in Wets (1966,1974a) and Parikh (1968).

The *simple recourse function* components $\Psi_i(\chi_i)$ can then be written as

$$\Psi_i(\chi_i) = q_i^+ \bar{\xi}_i - q_i^+ \chi_i + q_i \chi_i P_i(\chi_i) - q_i \int_{-\infty}^{\chi_i} \xi_i dP_i(\xi_i), \tag{2.2}$$

where $q_i = q_i^+ + q_i^- > 0$ and $\bar{\xi}_i = \int \xi_i dP_i(\xi_i)$. The functions $\chi \mapsto \Psi(\chi)$ and $\chi_i \mapsto \Psi_i(\chi_i)$ are continuous convex functions on $\Re^{m_2}$ and $\Re$, respectively (Wets (1974a)). The subdifferential at χ_i is given by

$$\partial\Psi_i(\chi_i) = \{\pi \mid -q_i^+ + q_i P_i^-(\chi_i) \leq \pi \leq -q_i^+ + q_i P_i(\chi_i)\}, \tag{2.3}$$

where $P_i^-(\chi_i) = \lim_{y \uparrow \chi_i} P_i(y)$ (Wets (1974a)). From (2.3), if P_i is continuous, then Ψ_i is differentiable.

The development of our sublinear approximation relies upon the use of conjugate functions. To find the conjugate of Ψ_i, first define

$$G_i(p) = \{y \mid P_i^-(y) \leq p \leq P_i(y)\}. \tag{2.4}$$

The conjugate of Ψ_i is written

$$\Psi_i^*(v_i) = \sup_{\chi_i}\{v_i \chi_i - \Psi_i(\chi_i)\}. \tag{2.5}$$

PROPOSITION 2.1. The conjugate function of Ψ_i defined in (2.2) is given by

$$\Psi_i^*(v_i) = \begin{cases} -q_i^+\overline{\xi}_i + (v_i + q_i^+)y - q_i y P_i(y) \\ +q_i \int_{-\infty}^{y} \xi_i dP_i(\xi_i), & \text{if } -q_i^+ < v_i < q_i^-, \\ -q_i^+\overline{\xi}_i & \text{if } v_i = -q_i^+, \\ q_i^-\overline{\xi}_i & \text{if } v_i = q_i^-, \\ +\infty, & \text{otherwise,} \end{cases} \tag{2.6}$$

where $y \in G_i(\frac{v_i+q_i^+}{q_i})$.

Proof. From (2.5) and (2.2), we have

$$\begin{aligned} \Psi_i^*(v_i) &= \sup_{\chi_i}\{v_i\chi_i - q_i^+\overline{\xi}_i + q_i^+\chi_i - q_i\chi_i P_i(\chi_i) \\ &\quad + q_i\int_{-\infty}^{\chi_i} \xi_i dP_i(\xi_i)\} \\ &= -q_i^+\overline{\xi}_i + \sup\{(v_i + q_i^+)\chi_i - q_i\chi_i P_i(\chi_i) \\ &\quad + q_i\int_{-\infty}^{\chi_i} \xi_i dP_i(\xi_i)\}. \end{aligned} \tag{2.7}$$

From (2.7), it is clear that $\Psi_i^*(v_i) = +\infty$ if $v_i < -q_i^+$ or $v_i > q_i^-$. We, therefore, assume that $-q_i^+ \le v_i \le q_i^-$. For $v_i \in (-q_i^+, q_i^-)$, the set of χ_i that attains the supremum in (2.7) is $\{\chi_i \mid P_i^-(\chi_i) \le \frac{v_i+q_i^+}{q_i} \le P_i(\chi_i)\} = G_i(\frac{v_i+q_i^+}{q_i})$. Substitution of $y \in G_i(\frac{v_i+q_i^+}{q_i})$ yields the first formula in (2.6). For v_i on the boundary of $[-q_i^+, q_i^-]$ and bounded distributions, the same argument applies. The formulas for unbounded distributions follow directly given our assumption of bounded first and second moments. ∎

The following corollaries follow immediately from Proposition 2.1.

COROLLARY 2.2. If P_i is continuous, then

$$\Psi_i^*(v_i) = \begin{cases} -q_i^+\overline{\xi}_i + q_i\int_{-\infty}^{y} \xi_i dP_i(\xi_i), & \text{if } -q_i^+ \le v_i \le q_i^-, \\ +\infty & \text{otherwise,} \end{cases} \tag{2.8}$$

where $y \in G_i(\frac{v_i+q_i^+}{q_i})$.

COROLLARY 2.3. If ξ_i has the degenerate distribution

$$P_i(\xi_i) = \begin{cases} 1 & \text{if } \xi_i \ge \overline{\xi}_i, \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

then

$$\Psi_i^*(v_i) = \begin{cases} v_i\bar{\xi}_i & \text{if } -q_i^+ \le v_i \le q_i^-, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.10}$$

In addition to Ψ_i^*, we are also interested in its subgradients.

PROPOSITION 2.4. The subdifferential of Ψ_i^* at $v_i \in$ ri (dom Ψ_i^*) (the relative interior of the effective domain of Ψ_i^*) is given by

$$\partial\Psi_i^*(v_i) = G_i\left(\frac{v_i + q_i^+}{q_i}\right). \tag{2.11}$$

Proof: The subgradients are found directly using (2.6). ∎

3. The Sublinear Upper Bound

Simple recourse functions are used here to approximate the general recourse function function ψ as defined in (1.2). We assume that $\psi(\chi, \xi)$ is finite for all values of χ and ξ, i.e., pos$W = \Re^{m_2}$. This assumption corresponds to ***complete recourse*** in stochastic programming. In practice, it can be achieved by introducing appropriate penalties in the recourse problem. In this case, the function ϕ defined by

$$\phi(\xi - \chi) = \psi(\chi, \xi) \tag{3.1}$$

is ***sublinear*** (positively homogeneous and convex). This property allows the simple recourse function approximation. The function ϕ is also ***polyhedral*** (i.e., its epigraph is a polyhedral cone).

Birge and Wets (1986a) introduced a method for approximating ψ by simple recourse functions. This method was based on solving the linear program:

$$\text{find } y \in \Re^{n_2} \text{ such that } Wy = e_i, y \ge 0, \tag{3.2}$$

$$\text{and } qy \text{ is minimized,}$$

where e_i is the i*th* unit m_2-vector. The optimal solution value of (3.2) is $q^+_{I(i)}$. If we substitute $-e_i$ for e_i, the optimal solution value is $-q^-_{I(i)}$. By sublinearity,

$$\psi(\chi, \xi) \le \psi_I(\chi, \xi) = \sum_{i=1}^{m_2} \psi_{I(i)}(\chi_i, \xi_i), \tag{3.3}$$

where

$$\psi_{I(i)}(\chi_i, \xi_i) = \begin{cases} q^+_{I(i)}(\xi_i - \chi_i), \xi_i \geq \chi_i \\ q^-_{I(i)}(\chi_i - \xi_i), \xi_i < \chi_i. \end{cases} \tag{3.4}$$

The function ψ_I is a *simple recourse function.*

By integration in (3.3), we have

$$\Psi(\chi) \leq \Psi_I(\chi),$$

where

$$\begin{aligned} \Psi_I(\chi) &= \int_{\Xi} \psi_I(\chi, \xi) dP(\xi) \\ &= \sum_{i=1}^{m_2} \{ \int_{\xi_i \geq \chi_i} q^+_{I(i)}(\xi_i - \chi_i) dP(\xi_i) \\ &+ \int_{\xi_i < \chi_i} q^-_{I(i)}(\chi_i - \xi_i) dP(\xi_i) \}. \end{aligned}$$

Note that Ψ_I is separable in the components of χ and only line integration is required in its computation.

Other approximations are obtained by considering directions other than $\pm e_i$ in (3.2). Let $h_1, ..., h_K \in \Re^{m_2}$ positively span $\Re^{m_2}$ (i.e.,$\Re^{m_2} =$ pos $[h_1, ..., h_K]$). Substitute $h_j, j = 1, ..., K$, for e_i in (3.2) and let the optimal solution values be $q_{H(j)}, j = 1, ..., K$. We then have

$$\psi(\chi, \xi) \leq \inf\{\sum_{j=1}^{K} \lambda_j q_{H(j)} \mid \sum_{j=1}^{K} \lambda_j h_j = \xi - \chi, \lambda_j \geq 0, j = 1, ..., K\}. \tag{3.5}$$

If $H = [h_1, ..., h_K]$ includes all columns of W, then (3.5) becomes an equality (see Birge and Wets (1986b)).

A difficulty in using (3.5) for approximating Ψ is that in general an optimization must still be performed inside the integral. This solution is immediate, however, if H is a *positive linear basis* for $\Re^{m_2}$, i.e., every point in $\Re^{m_2}$ corresponds to a unique positive combination of the h_i. A convenient choice for such a set is to use a linear basis $D = [d_1, ..., d_{m_2}]$ for $\Re^{m_2}$ and $-D = [-d_1, ..., -d_{m_2}]$, so that pos $[D, -D] = \Re^{m_2}$. In this case, $\xi - \chi =$

$\sum_{i=1}^{m_2}(D^{-1})_{i\cdot}(\xi-\chi)d_i$, where $(D^{-1})_{i\cdot}$ indicates the ith row of D^{-1}. The approximation is

$$\psi(\chi,\xi) \le \psi_D(\chi,\xi) = \sum_{i=1}^{m_2} \psi_{D(i)}(\chi,\xi) \tag{3.6}$$

where

$$\psi_{D(i)}(\chi,\xi) = \begin{cases} q^+_{D(i)}(D^{-1})_{i\cdot}(\xi-\chi) & \text{if } (D^{-1})_{i\cdot}(\xi-\chi) \ge 0, \\ q^-_{D(i)}(D^{-1})_{i\cdot}(\chi-\xi) & \text{if } (D^{-1})_{i\cdot}(\chi-\xi) > 0, \end{cases} \tag{3.7}$$

and $q^+_{D(i)}$ and $q^-_{D(i)}$ are the optimal solution values of (3.2) with d_i and $-d_i$, respectively, substituted for e_i and $-e_i$.

Several different bases are used in the sublinear approximation of ψ. For $\mathcal{D} = \{D^1,\ldots,D^L\}$, a set of linear bases for $\Re^{m_2}$,

$$\begin{aligned} \psi(\chi,\xi) &\le \inf\{\sum_{j=1}^{L}\lambda^j\psi_{D^j}(\chi,\xi^j) \mid \sum_{j=1}^{L}\lambda^j\xi^j = \xi, \sum_{j=1}^{L}\lambda^j = 1, \lambda^j \ge 0, \text{ for all } j\}, \\ &= \text{co}\,(\psi_{D^j}(\chi,\cdot), j=1,\ldots,L)(\xi), \end{aligned} \tag{3.8}$$

the function obtained by taking the convex hull of the epigraphs of $\psi_{D^j}, j=1,\ldots,L$. Of course, if $\mathcal{D}$ includes all linear bases in W, then (3.8) is satisfied as an equality. Although it might be simpler than (3.5), inequality (3.8) is still difficult to use computationally, again because of the minimization required inside the integral. A weaker, but usable inequality is obtained by reversing the infimum and integration. For this, we define

$$\Psi_D(\chi) = \sum_{i=1}^{m_2} \Psi_{D(i)}(\chi),$$

where

$$\begin{aligned} \Psi_{D(i)}(\chi) = &\int_{(D^{-1})_{i\cdot}(\xi-\chi)\ge 0} q^+_{D(i)}(D^{-1})_{i\cdot}(\xi-\chi)dP(\xi) \\ &+ \int_{(D^{-1})_{i\cdot}(\xi-\chi)<0} q^-_{D(i)}(D^{-1})_{i\cdot}(\chi-\xi)dP(\xi), \end{aligned} \tag{3.9}$$

PROPOSITION 3.1. Let $\mathcal{D}$ be a set of linear bases of $\Re^{m_2}$, then

$$\Psi(\chi) \le \text{co}\,\{\Psi_D, D\in\mathcal{D}\}(\chi). \tag{3.10}$$

Proof: Integration of (3.6) yields

$$\Psi(\chi) \leq \Psi_D(\chi) \tag{3.11}$$

for any $D \in \mathcal{D}$. So, $\sum_{i=1}^{m_2} \lambda^i \Psi(\chi^i) \leq \sum_{i=1}^{m_2} \lambda^i \Psi_D(\chi^i)$, for any $\lambda^i \geq 0$. Letting $\sum_{i=1}^{m_2} \lambda^i = 1$ and $\chi = \sum_{i=1}^{m_2} \lambda^i \chi^i$ yields (3.10) by the convexity of Ψ.∎

Equality in (3.10) can only be guaranteed in very special cases, even if $\mathcal{D}$ includes all linear bases from W. To see this simply observe that if $\mathcal{D}$ is rich enough, then

$$\int \text{co } \Psi_D = \int \inf_D \Psi_D \geq \inf_D \int \Psi_D$$

with strict inequality except in degenerate cases such as: $\mathcal{D}$ is a singleton, the Ψ_D's are linear, the probability measure is degenerate, etc. But as suggested by Proposition 3.1, co $\{\Psi_D, D \in \mathcal{D}\}(\chi)$ always provides us with an upper bound, that is relatively easy to compute, even when other procedures require more computations than can be implemented in realistic times (i.e., the solutions of more than 1000 linear programs for a single bound on problems with more than 10 random variables). We have also observed that the solution obtained from solving the "stochastic" program

$$\text{find } x \in \Re^{n_1}, \chi \in \Re^{m_2} \text{ such that } Ax = b, Tx = \chi, x \geq 0,$$

$$\text{and } z = cx + \text{ co } \{\Psi_D, D \in \mathcal{D}\}(\chi) \text{ is minimized,}$$

instead of (1.1), is usually a very good approximation of the optimal solution, much better than may be expected from the relatively lax inequality (3.10). Our experience shows that the function co $\{\Psi_D, D \in \mathcal{D}\}$ is "parallel" to Ψ, i.e.,

$$\partial\Psi(\chi) \approx \partial \text{ co } \{\Psi_D, D \in \mathcal{D}\}(\chi). \tag{3.12}$$

In the Appendix, we provide a heuristic argument and small example.

4. Dualization and Solution Procedures

Several questions must be answered in order to solve (1.1) with co$\{\Psi_D, D \in \mathcal{D}\}$ substituted for Ψ. The first concern is that finding the convex hull of a set of functions is itself

a difficult task. The second problem is to evaluate Ψ_D efficiently. A third area involves the choice of $\mathcal{D}$. We address each of these problems in this section. After substituting for Ψ, (1.1) becomes:

$$\text{find } x \in \Re^{n_1}, \chi \in \Re^{m_2} \text{ such that } Ax = b, Tx - \chi = 0, x \geq 0, \tag{4.1}$$

$$\text{and } z = cx + \text{ co } \{\Psi_D, D \in \mathcal{D}\}(\chi) \text{ is minimized.}$$

Instead of solving (4.1), we consider a dual program to (4.1). The dual program has a computational advantage because the convex hull operation is replaced by a supremum.

PROPOSITION 4.1. A dual program to (4.1) is given by

$$\text{find } \sigma \in \Re^{m_1}, \pi \in \Re^{m_2} \text{ such that } \sigma A + \pi T \leq c \tag{4.2}$$

$$\text{and } w = \sigma b - (\sup_{D \in \mathcal{D}} \Psi_D^*)(-\pi) \text{ is maximized,}$$

where Ψ_D^* is the conjugate function of Ψ and where the optimal value of (4.2), $w^* = z^*$, the optimal value of (4.1).

Proof: A general dual of (4.1) (see, e.g., Rockafellar (1974), Geoffrion (1971)) can be written as

$$\max_{\pi,\sigma}\{\inf_{x \geq 0, \chi} cx + g(\chi) + \sigma(b - Ax) + \pi(\chi - Tx)\} \tag{4.3}$$

where $g(\chi) = \text{co } \{\Psi_D \mid D \in \mathcal{D}\}(\chi)$. Since (4.1) involves linear constraints, the optimal values of (4.3) and (4.1) are the same. We can rewrite (4.3) as

$$\max_{\pi,\sigma}\{\inf_{x \geq 0, \chi} (c - \sigma A - \pi T)x - (-\pi\chi - g(\chi)) + \sigma b\}, \tag{4.4}$$

which is equivalent to

$$\text{find } \sigma \in \Re^{m_1}, \pi \in \Re^{m_2} \text{ such that } \sigma A + \pi T \leq c \tag{4.5}$$

$$\text{and } w = \sigma b - g^*(-\pi) \text{ is maximized.}$$

By Theorem 16.5 of Rockafellar (1970),

$$(\text{ co } \{\Psi_D \mid D \in \mathcal{D}\}(\chi))^* = \sup\{\Psi_D^* \mid D \in \mathcal{D}\}, \tag{4.6}$$

yielding (4.2). ∎

A solution of the dual program (4.2) is simpler to compute than a solution of (4.1) because $g^*(-\pi)$ is much easier to evaluate than $g(\chi)$. We must, however, obtain an expression for Ψ_D^*. First, let $P_{D(i)}$ be the distribution function of $\zeta_i = (D^{-1})_{i\cdot}(\xi - \chi)$ and let $\bar{\zeta}_i = \int \zeta_i dP_{D(i)}(\zeta_i)$. Also define

$$G_{D(i)}(p) = \{y \mid P^-_{D(i)}(y) \leq p \leq P_{D(i)}(y)\}.$$

PROPOSITION 4.2. The conjugate function of Ψ_D is given by

$$\Psi_D^*(v) = \sum_{i=1}^{m_2} \Psi^*_{D(i)}(v), \tag{4.7}$$

where

$$\Psi^*_{D(i)}(v) = \begin{cases} -q^+_{D(i)}\bar{\zeta}_i + (vD_{\cdot i} + q^+_{D(i)})y - q_{D(i)}yP_{D(i)}(y) & \\ \quad + q_{D(i)}\int_{-\infty}^{y} \zeta_i dP_{D(i)}(\zeta_i), & \text{if } -q^+_{D(i)} < vD_{\cdot i} < q^-_{D(i)}, \\ -q^+_{D(i)}\bar{\zeta}_i & \text{if } vD_{\cdot i} = -q^+_{D(i)}, \\ q^-_{D(i)}\bar{\zeta}_i & \text{if } vD_{\cdot i} = q^-_{D(i)}, \\ +\infty & \text{otherwise,} \end{cases} \tag{4.8}$$

where $q_{D(i)} = q^+_{D(i)} + q^-_{D(i)}$ and $y \in G_{D(i)}(\frac{vD_{\cdot i} + q^+_{D(i)}}{q_{D(i)}})$.

Proof: From (3.9), observe that

$$\Psi_{D(i)}(\chi) = \int \psi_i^s((D^{-1}\chi)_i, \zeta_i) dP_{D(i)}(\zeta_i), \tag{4.9}$$

where ψ_i^s is a simple recourse function as in (2.1) with $q^+_{D(i)}$ and $q^-_{D(i)}$ substituted for q_i^+ and q_i^- respectively. From (4.9), it follows that

$$\Psi_D(\chi) = \sum_{i=1}^{m_2} \Psi_i^s((D^{-1}\chi)_i) = \Psi^s(D^{-1}\chi), \tag{4.10}$$

where Ψ_i^s has the form of the simple recourse function in (2.2) and Ψ^s is the simple recourse function defined for the random variables $\varsigma_1, ..., \varsigma_{m_2}$ in place of $\xi_1, ..., \xi_{m_2}$.

The dual of Ψ_D is then (see Theorem 16.3 Rockafellar (1970))

$$\Psi_D^* = (\Psi^s D^{-1})^* = D^{-1}(\Psi^s)^*, \tag{4.11}$$

where

$$D^{-1}(\Psi^s)^*(v) = (\Psi^s)^*(vD). \tag{4.12}$$

Applying (4.11) and (4.12) to (2.6) yields the result. ∎

A subgradient of Ψ_D^* is used in the optimization procedure. It is also calculated from the subdifferential of $(\Psi^s)^*$.

PROPOSITION 4.3. The subdifferential of Ψ_D^* at $v \in$ ri (dom Ψ_D^*) is

$$\partial(\Psi_D^*)(v) = \{Dy \mid y \in \Re^{m_2}, y_i \in G_{D(i)}(\frac{vD_{\cdot i} + q_{D(i)}^+}{q_{D(i)}}), i = 1, ..., m_2\}. \tag{4.13}$$

Proof: If w^* is a subgradient of $(\Psi^s)^*$ at w, then

$$(z - w) \cdot w^* + (\Psi^s)^*(w) \leq (\Psi^s)^*(z). \tag{4.14}$$

From (4.11) and (4.12) if $z = uD$ and $w = vD$, (4.14) becomes

$$(v - u)Dw^* + \Psi_D^*(v) \leq \Psi_D^*(u). \tag{4.15}$$

From Proposition 2.4, $w^* = (w_1^*, ..., w_{m_2}^*)$ where $w_i^* \in G_i(\frac{vD_{\cdot i} + q_i^{D+}}{q_i^D})$, so Dy in (4.13) is a subgradient of Ψ_D^* at v. A reverse argument shows that any subgradient of Ψ_D^* has the form Dy in (4.12), proving the result. ∎

The expressions for Ψ_D^* and $\partial\Psi_D^*$ in (4.7–8) and (4.13) are used in an optimization procedure for (4.2). A difficulty is that, even when each Ψ_D^* is differentiable (i.e., the distribution function of ς_i is strictly increasing on its support, but not necessarily continuous), the objective function in (4.2) is not necessarily differentiable. Nondifferentiable methods (see, e.g., Lemaréchal (1978), Lemaréchal et al. (1981), Nazareth and Wets (1986), Polak (1987), Wolfe (1975)) can be applied to this program. We, however, transform (4.2) into

a smooth optimization problem with nonlinear constraints (as suggested, for example, in Gill, Murray, and Wright (1981)). The new problem becomes

$$\text{find } \sigma \in \Re^{m_1}, \pi \in \Re^{m_2}, \theta \in \Re \text{ such that} \tag{4.16}$$

$$\begin{array}{lllll} \sigma A & +\pi T & & \leq c & \\ & \Psi_D^*(-\pi) & -\theta & \leq 0, & \text{for all } D \in \mathcal{D}, \text{ and} \\ w = \sigma b & & -\theta & & \text{is maximized.} \end{array}$$

General methods for optimization problems with nonlinear constraints can be applied to (4.16). In our examples, we use the MINOS computer code (Murtagh and Saunders (1980)) to solve (4.16). Note that (4.16) is similar in form to a master problem in a *linear* outer approximation algorithm (e.g., the L-shaped method), but we solve the dual and use a *nonlinear* outer approximation.

5. Implementation Considerations

In solving (4.2) or (4.16) one needs to find $P_{D(i)}$, the distribution function of ς_i, and $G_{D(i)}$, the inverse function. If the random variables ξ_j are independently, normally distributed with means, μ_j, and variances, σ_j^2, then $\varsigma_i = (D^{-1})_{i\cdot}(\xi - \chi)$ is also normally distributed with mean $\sum_{j=1}^{m_2}(D^{-1})_{ij}(\mu_j - \chi_j)$ and variance $\sum_{j=1}^{m_2}(D^{-1})_{ij}^2\sigma_j^2$. (Note that degenerate and correlated random variables can also be included among the ξ_i with ς_i remaining normally distributed.) Other distributions require special schemes in order to integrate with respect to ς_i. In our experiments, we used normally distributed random variables because of the ease in performing these computations. Upper bounds can be obtained for other distributions by using the approaches in Birge and Wets (1986a, 1987).

Given a problem form ((4.2) or (4.16)) and a method for finding $P_{D(i)}$, the set of bases must still be chosen. For the problem of approximating Ψ, it appears that the matrices to include in $\mathcal{D}$ should be chosen so that the level sets of ψ_D cover high probability regions of the level sets of ψ. This coverage, however, depends on χ, so the choices should be good for a range of values of χ (that would ideally include the optimal value of χ). In our experience, the identity provided a good starting basis , especially when the optimal χ was close to $\bar{\xi}$.

We implemented several basis generation procedures that started with the identity as the first basis and then included additional bases. Generating random bases from the set of all bases proved inefficient because the corresponding functions Ψ_D^* often did not improve the solution. Bases were then generated from the set of optimal bases for some ξ by solving (1.2) for varying values of ξ and χ. A patterned choices of values for ξ (using $\bar{\xi}_i \pm 3\sigma_i$ for each i) proved slightly more efficient than random selection of ξ_i and was, therefore, used in the experiments described below. Two different choices for χ were implemented. On the kth solution of (4.16) with new bases added to $\mathcal{D}$, we used $\chi = \chi^{k-1}$ or $-\chi^{k-1}$ where χ^{k-1} was the optimal value of χ from the $(k-1)$th solution of (4.16). In our experiments, $\chi = -\chi^{k-1}$ proved more effective because it included a broader class of bases by exploring different regions of χ.

The sublinear upper bounding method with this implementation is given below. The basic tolerance parameters are $TOLINF$ for infinity and $TOLCONV$ for convergence.

Sublinear Upper Bounding Method

Step 0. Initialization. Let $D^1 = I$, $K = 1$, $\mathcal{D}=\{D^1, \cdots, D^K\}$, $J = \emptyset$, $w_{old} = TOLINF$, $l = 0$, $\chi = 0$.

Step 1. Add New Basis. Let $D^K = [D^K(1), \ldots, D^K(m_2)]$.

a. For $i = 1, \ldots, m_2$, if $D^K(i) \notin J$, let $J = J \cup D^K(i)$ and find $q^+_{D^k(i)}, q^-_{D^k(i)}$.

b. For $i = 1, \ldots, m_2$, calculate parameters for $P_{D(i)}$.

If $K = 1$, go to 3.

Step 2. Search for New Basis. For $i = 1, \ldots, m_2$, $l = \sum_{i=1}^{m_2}(k(i)+1)3^{i-1}$, let $h(i) = \bar{\xi}_i + 3k(i)\sigma_i$. Find $\psi(\chi, h)$ with an optimal basis, D'.

If $D' \notin \mathcal{D}$, $D^{K+1} = D'$, $K = K + 1$, go to 1.

Else, if $l < 3^{m_2}$, $l = l + 1$, repeat.

Else, go to 3.

Step 3. Find New Bound. Solve (4.16) to obtain w_{new} and let the dual variables associated with the linear constraints be x. Let $\chi = Tx$.

If $TOLCONV < (w_{old} - w_{new})$ $(/w_{new}$ if $w_{new} \neq 0)$, go to 2.

Else, stop, w_{new} is the sublinear upper bound.

6. Numerical Results

Formulation 4.16 was used with the MINOS/AUGMENTED (MINOS Version 4.0) computer program for nonlinearly constrained problems. This implementation on The University of Michigan's Amdahl 5860 Computer found optimal solutions for (4.16) for all of the test problems tried.

The appropriate use of tolerances on constraint satisfaction was especially important in our implementation. A value of 10^{-5} was used for ROW TOLERANCE in the MINOS SPECS file. This allowed some flexibility in satisfying the constraints without creating large infeasibilities. To avoid infinite values of Ψ_D^*, the constraints

$$-q_{D(i)}^{+} \leq vD_i \leq q_{D(i)}^{-}, \tag{5.1}$$

were added to (4.16). A penalty term was included in the subgradient definition from (4.13) when (5.1) was satisfied as an equality (within the tolerance). This made the subgradient definition consistent at the boundary of the effective domain of Ψ_D^*.

Problem testing initially involved simple recourse problems which were solved exactly by the formulation in (4.16). After conducting this check, the method was applied to general recourse problems. The results reported here apply to the small energy decision problem in Louveaux (1987). In this problem, $m_1 = 2$, $n_1 = 4$, $m_2 = 7$, and $n_2 = 12$. Of the seven recourse problem constraints, four are balancing constraints that are fixed at zero and three are demand constraints that are stochastic. The demands were assumed independent and normally distributed. Different problems were generated by varying the means and standard deviations of these random variables (see Table 1).

To check the accuracy of the sublinear approximation method, the test problems were also solved using the L-shaped code, NDSP(Birge (1985)), with upper bounds from the Edmundson-Madansky inequality (Madansky (1959)) and lower bounds from Jensen's inequality. After the kth solution of the problem with these bounds, the bounds were refined

for each ξ_i until NDSP's limit of 125 realizations of the random vector ξ was reached. This occurs here at $k = 4$, but note that the Edmundson–Madansky bound in NDSP cannot be applied even once to provide an upper bound for problems with more than seven random variables. The sublinear upper bound does not have this limitation.

The results for the upper bounds and CPU second times for each iteration appear in Table 2 under the NDSP columns. The upper and lower bounds on the final iteration value appear in Table 3. Note that in Problem 8 after 33.80 CPU seconds, NDSP terminated because a limit of 100 L-shaped algorithm iterations (corresponding to 100 cutting planes) had been performed. Execution was not continued because of large row residuals caused by instability in the bases including these cuts (see Birge (1986) for a discussion of this phenomenon).

The sublinear upper bounding approximation results appear in Tables 2 and 3. As given in Section 5, the algorithm is terminated when the addition of new bases does not change the solution. This is indicated by "b" in Table 2. Further improvement would be possible using groups of random variables as noted in the next section. The upper bound result of solving (4.16) and the CPU seconds to obtain this result appear in Table 2. The computation times include some iteration logging that is comparable but not identical to iteration log times included in the CPU times for NDSP.

Table 3 provides bounds on the objective values $Z(NDSP) = cx_{ND} + \Psi(\chi_{ND})$ and $Z(SL) = cx_{SL} + \Psi(\chi_{SL})$, where (x_{ND}, χ_{ND}) is the last solution found by NDSP and (x_{SL}, χ_{SL}) is the last solution found by the sublinear upper bounding (SL) method. Upper bounds are found using the Edmundson-Madansky inequality, and lower bounds are found using the Jensen inequality. The relative differences between the means of the upper and lower bounds given by NDSP and the sublinear method are also given in Table 3.

From Table 2, it appears that the sublinear method provides upper bounds that are close to the NDSP values within times that are comparable with the NDSP times. Table 3 indicates how well the x and χ solutions from the sublinear method perform in (1.1). The upper and lower bounds result from upper and lower approximations of Ψ. In Table 3, note that the mean values of the upper and lower bounds on $Z(SL)$ are at most 5.8%

worse than the NDSP values and are at best 12.5% better.

A key property to note in examining Table 2 is that the sublinear method times are comparable for all problems, but the accuracy relative to the NDSP bounds does not degrade as problem size increases. The NDSP approximations, however, become much worse as additional random variables are included into the problem. The true advantage of the sublinear method is, therefore, in larger problems, as is evident in Problems 7–9. Here, the sublinear method obtains relatively good bounds (within 25% of the best lower bound) in fractions of the times for NDSP. These experiments were limited to problems where an Edmundson–Madansky could be obtained. Again, it should be emphasized that the sublinear upper bound can be applied to problems where the Edmundson–Madansky would not be calculable.

Table 1. Problem Parameters

Problem	Random Variables Parameters					
	μ_1	σ_1	μ_2	σ_2	μ_3	σ_3
1	5.0	1.0	4.0	0.0	3.0	0.0
2	5.0	0.0	4.0	1.0	3.0	0.0
3	5.0	0.0	4.0	0.0	3.0	1.0
4	5.0	1.0	4.0	1.0	3.0	0.0
5	5.0	1.0	4.0	0.0	3.0	1.0
6	5.0	0.0	4.0	1.0	3.0	1.0
7	5.0	1.0	4.0	1.0	3.0	1.0
8	2.0	1.0	4.0	1.0	3.0	1.5
9	4.0	1.0	5.0	1.0	5.0	1.0

Table 2. Results on Upper Bounds

Problem	Iteration	NDSP		Sublinear Method	
		Bound	CPUs[a]	Bound	CPUs
1	1	395.	0.41	452.	0.82
	2	367.	0.75	385.	1.79
	3	366.	1.12	383.	2.38
	4	366.	1.27	b	2.75
2	1	345.	0.38	398.	0.54
	2	335.	0.71	345.	1.48
	3	335.	0.79	b	1.71
	4	335.	0.88	–	–
3	1	291.	0.42	304.	1.44
	2	289.	0.47	289.	2.03
	3	b	–	b	2.42
	4	–	–	–	–
4	1	1180.	0.42	572.	3.37
	2	454.	1.01	440.	5.78
	3	425.	1.85	b	6.22
	4	424.	2.88	–	–
5	1	1140.	0.50	478.	0.63
	2	505.	1.19	407.	2.14
	3	388.	1.97	402.	3.82
	4	377.	4.30	b	4.37
6	1	1100.	0.37	425.	2.17
	2	473.	0.73	350.	7.01
	3	357.	1.05	b	7.53
	4	347.	1.54	–	–
7	1	2060.	0.51	617.	0.56
	2	1100.	2.09	470.	3.56
	3	488.	9.59	b	4.09
	4	466.	26.44	–	–
8	1	2360.	0.61	511.	0.62
	2	1350.	2.17	359.	2.82
	3	491.	11.63	b	3.52
	4	c	33.80	–	–
9	1	2420.	0.48	626.	0.72
	2	878.	1.49	476.	5.07
	3	475.	6.94	b	5.62
	4	471.	13.52	–	–

a — CPU seconds

b — No improvement

c — Iteration limit exceeded

Table 3. Upper and Lower Bounds on Last Solution Value

Problem	NDSP		Sublinear Approximation		
	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Mean Difference*
1	366.	363.	371.	370.	.016
2	335.	334.	343.	343.	.026
3	289.	289.	289.	289.	.000
4	424.	419.	438.	436.	.037
5	377.	374.	390.	389.	.037
6	347.	345.	350.	349.	.010
7	466.	438.	494.	462.	.058
8	491.	316.	369.	337.	-.125
9	471.	433.	486.	465.	.052

* — $((UB(SL) + LB(SL))/2 - (UB(NDSP) + LB(NDSP))/2)/((UB(NDSP) + LB(NDSP))/2)$

7. Discussion and Extensions

The results in Section 6 indicate that the sublinear upper bounding method is a good upper bounding, approximation procedure for stochastic programs, especially when other bounds are not computationally tractable. The sublinear method provides good bounds on the optimal objective value quickly. The solutions obtained from it are also generally close in value to the solutions obtained by other solution procedures. In addition, the sublinear method can be applied to problems with many random variables where other methods cannot be applied. This ability for general stochastic programs sets the sublinear method apart from all other methods. The sublinear bound, however, only majorizes the optimal objective value. Lower bounds, obtained from the Jensen inequality or other outer linearization approach (see Marti [1975] and Birge and Wets [1986a]) should be used in a full optimization procedure.

A convergent procedure for bounding the recourse function value can be obtained by evaluating the recourse function with respect to groups of random variables. As the number of random variable in these groups increases the bound is tightened until, if all random variables are included in a single integration, the true objective value is obtained. This procedure is, of course, only useful as long as the integrals are simple enough to allow easy calculation. For example, suppose $m_2 = 3$,

$$\begin{aligned} \psi_2(\chi, \xi) &= \min\{qy \mid Wy = (\xi_1 - \chi_1, \xi_2 - \chi_2, 0)^T, y \geq 0\} \\ &= \pi_1^i \varsigma_1 + \pi_2^i \varsigma_2 \end{aligned} \tag{7.1}$$

where $(\varsigma_1, \varsigma_2) = (\xi_1 - \chi_1, \xi_2 - \chi_2) \in R^i, i = 1, ..., r$, and we can evaluate

$$\Psi_2(\chi_1, \chi_2) = \sum_{i=1}^{r} \int_{R^i} (\pi_1^i \varsigma_1 + \pi_2^i \varsigma_2) dP(\varsigma_1, \varsigma_2). \tag{7.2}$$

We then combine the approximation in (7.2) with a simple recourse function approximation, $\Psi_3(\chi_3)$, evaluated with respect to $\varsigma_3 = \xi_3 - \chi_3$. The result is another upper bound on Ψ, where

$$\Psi(\chi) \leq \Psi_2(\chi_1, \chi_2) + \Psi_3(\chi_3). \tag{7.3}$$

Again, convex combinations of approximations of the form (7.3) for every pair of random variables can be used to approximate Ψ from above. The practicality of this approach of course relies upon the possibility to evaluate integrals as in (7.2) efficiently. For large groups of random variables, this effort may negate any advantages from this approximation scheme.

Appendix

The following heuristic argument and small example give some justification for the observation in (3.12). Suppose (x^0, χ^0) is a feasible solution of (1.1) that is near the optimal solution. Suppose $\mathcal{D}$ contains all bases (in W) that correspond to basic optimal solutions, for any possible pair (ξ, χ), of the linear program that defines $\psi(\chi, \xi)$, cf. Wets (1974b) for the Basis Decomposition Theorem. Also, suppose that for some $D_0 \in \mathcal{D}$,

$$\text{co } \{\Psi_D, D \in \mathcal{D}\}(\chi^0) = \Psi_{D_0}(\chi^0). \tag{A.1}$$

(A convex combination of Ψ_D values would, in general, appear above. We consider a single function to simplify the argument.) Given (A.1)

$$\begin{aligned} \partial\{ \text{ co } \Psi_D\}(\chi^0) &= \partial\Psi_{D_0}(\chi^0) \\ &= \int \partial\psi_{D_0}(\chi^0, \xi) dP(\xi). \end{aligned}$$

Observe that $\partial\psi_{D_0}(\chi^0, \xi) = -q^+_{D_0}$ if $\xi \in \Xi_0$, where

$$\Xi_0 = \{\xi \mid D_0^{-1}\xi \geq D_0^{-1}\chi^0\}.$$

Since Ψ_{D_0} gives the convex hull operation value, the region of Ξ in which Ψ_{D_0} is exact should be large. One can then reasonably expect Ξ_0, or more generally the region surrounding Ξ_0, to contain most of the probability mass of Ξ. Hence the subgradients of $\partial\Psi_{D_0}$ at χ^0 are approximated by $-q^+_{D_0}$. On the other hand (Section 7, Wets (1974b)), excluding possibly some boundary cases,

$$\begin{aligned} \partial\Psi(\chi^0) &= \int \partial\psi(\chi^0, \xi) dP(\xi) \\ &= \sum_{D \in \mathcal{D}} (-\pi_D) P[D^{-1}\xi \geq D^{-1}\chi^0] \end{aligned} \tag{A.2}$$

where π_D are the multipliers associated with the basis D. Again, since Ξ_0 and its neighbors occupy much of the probability mass of Ξ, $\partial\Psi(\chi^0)$ is reasonably well approximated by $-\pi_{D_0}$. From construction, $q^+_{D_0} = \pi_{D_0}$. Hence, both ∂ co $\{\Psi_D\}$ and $\partial\Psi$ at χ^0 are near $-\pi_{D_0}$.

As an example, consider the stochastic program:

$$\text{find } x, y_1^+, y_1^-, y_2^+, y_2^-, y_3 \geq 0 \text{ such that} \tag{A.3}$$

$$\begin{array}{lllllll} x & & & & & & \leq 1 \\ t_1 x & +y_1^+ & -y_1^- & & & +y_3 & = \xi_1 \\ t_2 x & & & +y_2^+ & -y_2^- & +y_3 & = \xi_2 \end{array}$$

$$\text{to minimize } z = cx + \mathrm{E}_\xi[y_1^+ + y_1^- + y_2^+ + y_2^- + y_3].$$

where ξ_i is uniformly distributed on $[0,1]$, for $i = 1,2$. The optimal basis (or the related basis with positive coordinates) is indicated for each region of $[0,1] \times [0,1]$ in Figure 1. The three bases are

$$D^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, D^2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, D^3 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

We use $D^1,, D^2, D^3$ to develop approximating functions Ψ_{D^1}, Ψ_{D^2} and Ψ_{D^3}.

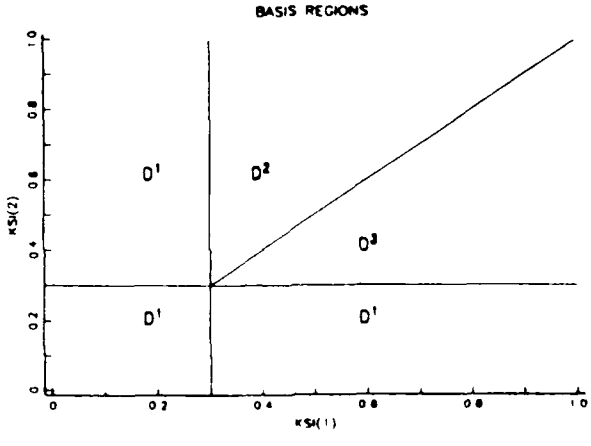


Figure 1. Optimal Subproblem Basis Regions

A comparison of the three approximations and the exact value of Ψ is given in Figure 2, where $\chi_2 = 0.1$. Note that the slope of the convex hull of the approximating functions

closely approximates the slope of Ψ. At low values of χ_1, the D^3 and D^2 optimal region occupy most of Ξ and hence, convex combinations of their subgradients provide good approximations of $\partial\Psi$. At higher values of χ_1, the optimal regions for D^3 and D^2 diminish, so that $\partial\Psi_{D^1}$ provides a good approximation of $\partial\Psi$.

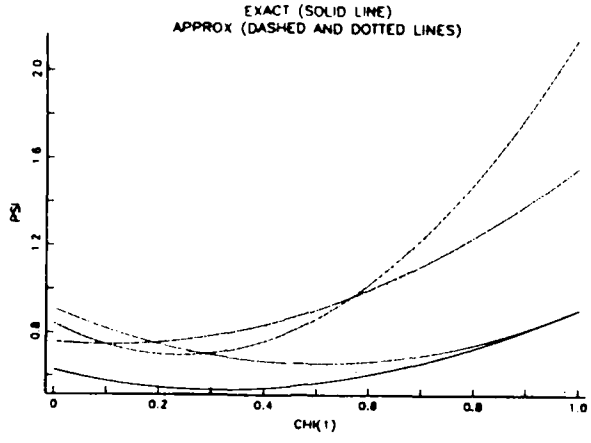


Figure 2. Comparison of ψ Values.

8. References

J.R. Birge, "Decomposition and partitioning methods for multi-stage stochastic linear programs," *Operations Research* 33 (1985) 989–1007.

J.R. Birge,"An L–shaped computer code for multi–stage stochastic linear programs," in: Y. Ermoliev and R. Wets, eds., *Numerical Methods in Stochastic Optimization* (Springer–Verlag, Berlin,1987) to appear.

J.R. Birge and Roger J–B. Wets, "Approximations and error bounds in stochastic programming,"in: Y. Tong, ed., *Inequalities in Statistics and Probability*, IMS Lecture Notes–Monograph Series Vol. 5 (Institute of Mathematical Statistics, Hayward, CA, 1984) pp. 178–186.

J.R. Birge and Roger J–B. Wets, "Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse," *Mathematical Programming Study* 27 (1986) 54–102.

J.R. Birge and R. J–B. Wets, "On–line solution of linear programs using sublinear functions," Technical Report No. 86–25, Department of Industrial and Operations Engineering, The University of Michigan (Ann Arbor, MI, 1986).

J.R. Birge and R. J–B. Wets, "Computing bounds for stochastic programming problems by means of a generalized moment problem," *Mathematics of Operations Research* 12 (1987) 149–162.

G.B. Dantzig and A. Madansky, "On the solution of two–stage linear programs under uncertainty," in: J. Neyman, ed., *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1961) pp. 165–176.

Y. Ermoliev, "Stochastic quasigradient methods and their applications to systems optimization," *Stochastics* 9 (1983) 1–36.

Y. Ermoliev and A. Gaivoronski, "Stochastic quasigradient methods and their implementation," in: Y. Ermoliev and R. Wets, eds., *Numerical Methods in Stochastic Optimization* (Springer–Verlag, Berlin, 1987) to appear.

A. Geoffrion, "Duality in nonlinear programming: a simplified applications-oriented development," *SIAM Review* 13 (1971) 1–37.

P.E. Gill, W. Murray and M.H. Wright, *Practical optimization* (Academic Press, London and New York,1981).

C.C. Huang, W. Ziemba, and A. Ben-Tal, "Bounds on the expectation of a convex function of a random variable: with applications to stochastic programming," *Operations Research* 25 (1979) 315–325.

P. Kall, "Computational methods for solving two–stage stochastic linear programming problems," *Z. Angew. Math. Phys.* 30 (1979) 261–271.

P. Kall, K. Frauendorfer and A. Ruszczynski, "Approximation techniques in stochastic programming," in: Y. Ermoliev and R. Wets, eds., *Numerical Methods in Stochastic Optimization* (Springer–Verlag, Berlin, 1987) to appear.

P. Kall and E. Keller, "Computational experience in solving stochastic programs," Technical Report, Institut für Operations Research, Universität Zürich, (Zürich, 1983).

P. Kall and D. Stoyan, "Solving stochastic programming problems with recourse including error bounds," *Math. Operationsforsch. Statist. Ser. Optim.* 13 (1982), 431–447.

C. Lemaréchal, "Bundle methods in nondifferentiable optimization," in: C. Lemaréchal and R. Mifflin, eds., *Nonsmooth Optimization*, (Pergamon Press, Oxford, 1978) pp. 79–102.

C. Lemaréchal, J.J. Strodiot and A. Bihain, "On a bundle algorithm for nonsmooth optimization," in: O. Mangasarian, S. Robinson and R. Meyer, eds., *Nonlinear Programming 4* (Academic Press, New York, 1981) pp. 245–282.

F.V. Louveaux, "Optimal investment for electricity generation: a stochastic model and a test problem," in: Y. Ermoliev and R. Wets, eds., *Numerical Methods in Stochastic Optimization* (Springer–Verlag, Berlin, 1987) to appear.

A. Madansky, "Bounds on the expectation of a convex function of a multivariate

random variable," *Ann. Math. Statist.* 30 (1959) 743–746.

K. Marti, "Approximationen von Entscheidungsproblemen mit linearer Ergebnisfunktion und positiv homogener, subadditiver Verlusfunktion," *Z. Wahrsch. Verw. Gebiete* 31 (1975) 203–233.

B.A. Murtagh and M.A. Saunders, "MINOS/AUGMENTED user's manual," Technical Report SOL 80–14, Systems Optimization Laboratory, Department of Operations Research, Stanford University (Stanford, CA, 1980).

J.L. Nazareth and R. J–B. Wets, "Algorithms for stochastic programs: the case of nonstochastic tenders," *Mathematical Programming Study* 28 (1986) 1–28.

S.C. Parikh, "Lecture notes on stochastic programming", Unpublished lecture notes, University of California (Berkeley, CA, 1968).

E. Polak,"On the mathematical foundations of nondifferentiable optimization in engineering design," *SIAM Review* 29 (1987) 21–90.

L. Qi, "An alternating method for stochastic linear programming with simple recourse," *Mathematical Programming Study* 27 (1986) 183–190.

R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, New Jersey, 1970).

R.T. Rockafellar, *Conjugate duality and optimization* (SIAM Publications, Philadelphia, 1974).

R.T. Rockafellar, *Network flows and monotropic optimization* (Wiley, New York, 1984).

B. Strazicky, "Computational experience with an algorithm for discrete recourse problems," in: M.A.H. Dempster, ed., *Stochastic Programming, Proceedings of the 1974 Oxford International Conference* (Academic Press, London, 1980) pp. 263–274.

R. Van Slyke and R. J–B. Wets, "L–shaped linear programs with application to optimal control and stochastic programming," *SIAM Journal Appl. Math.* 17 (1969) 638–663.

S.W. Wallace, "A piecewise linear upper bound on the network recourse problem,"

Mathematical Programming (1987) to appear.

R. J-B. Wets, "Programming under uncertainty: the complete problem," *Z. Wahrsch. Verw. Gebiete* 4 (1966) 316–339.

R. J-B. Wets, "Stochastic programming," Unpublished lecture notes, University of Kentucky (Lexington, KY, 1974).

R. J-B. Wets, "Stochastic programs with fixed recourse: the equivalent deterministic problem," *SIAM Review* 16 (1974) 309–339.

R. J-B. Wets, "Solving stochastic programs with simple recourse, II," in: *Proceedings of 1975 Conference on Information Sciences and Systems* (John Hopkins University, Baltimore, Maryland, 1975).

R. J-B. Wets, "Large scale linear programming techniques in stochastic programming," in: Y. Ermoliev and R. Wets, eds., *Numerical Methods in Stochastic Optimization* (Springer–Verlag, Berlin, 1987) to appear.

P. Wolfe, "A method of conjugate subgradients for minimizing convex functions," *Mathematical Programming Study* 3 (1975) 145–173.

END

DATE

FILMED

DEC.

1987